韓國 컴퓨터情報學會 論文誌
第12卷 第6號, 2007. 12.

2007-12-6-1-10

변형 커버리지 함수를 고려한 ENHPP 소프트웨어 신뢰성장 모형에 관한 비교 연구

김 희 철*, 김 평 구**

The Comparative Study for ENHPP Software Reliability Growth Model based on Modified Coverage Function

Kim Hee Cheul *, Kim Pyong Koo **

요 약

유한고장수를 가진 비동질적인 포아송 과정에 기초한 모형들에서 잔존 오류 1개당 고장 발생률은 일반적으로 상수, 혹은 단조증가 및 단조 감소 추세를 가지고 있다. 소프트웨어 제품의 정확한 인도시기를 예측하거나 효용성 및 신뢰성을 예측하기 위해서는 소프트웨어 테스팅 과정에서 중요한 요소인 테스트 커버리지를 이용하면 보다 효율적인 테스팅 작업을 할 수 있다. 본 논문에서는 기존의 소프트웨어 신뢰성 모형인 지수 커버리지 모형과 S-커버리지 모형을 적용하고 이 분야에 적용 될 수 있는 변형 커버리지 모형(중첩모형 및 혼합모형) 비교 문제를 제안하였다. 고장 간격시간으로 구성된 자료를 이용한 모수추정 방법은 최우추정법 과 수치해석 방법인 이분법을 사용하여 모수 추정을 실시하고 효율적인 모형 선택은 편차자승합(SSE)을 이용하였다.

Abstract

Finite failure NHPP models presented in the literature exhibit either constant, monotonic increasing or monotonic decreasing failure occurrence rates per fault. Accurate predictions of software release times, and estimation of the reliability and availability of a software product require quality of a critical element of the software testing process : test coverage. This model called Enhanced non-homogeneous Poission process(ENHPP). In this paper, exponential coverage and S-type model was reviewed, proposes modified(the superposition and mixture) model, which make out efficiency application for software reliability. Algorithm to estimate the parameters used to maximum likelihood estimator and bisection method, model selection based on SSE statistics for the sake of efficient model, was employed.

▸ Keyword : Test Coverage, ENHPP, Superposition Model, Mixture model , SSE

• 제1저자 : 김희철
• 접수일 : 2007.10.26, 심사일 : 2007.11.12, 심사완료일 : 2007.12.15.
* 남서울대학교 산업경영공학과, ** 충청대학 품질경영과

Ⅰ. 서 론

소프트웨어 테스팅 단계에서 소프트웨어 고장 수(Number of failure)와 고장간격시간에 의해 소프트웨어 고장현상을 수리적으로 접근 하면 소프트웨어에 대한 평가를 보다 쉽게 할 수 있으며 신뢰도 모형에 의해 소프트웨어 고장 수, 소프트웨어 고장발생간격시간, 소프트웨어 신뢰도 및 고장률 등의 신뢰성 평가측도들이 추정되어 미래의 고장시간을 예측할 수도 있다. 신뢰도에서 관측시간에 발견된 고장수를 모형화 하는데 비동질적 포아송 과정(Non-homogeneous Poission process; NHPP)이 널리 사용하여 왔다. 이러한 NHPP모형은 강도함수(Intensity function)와 평균값함수(Mean value function)에 의존한다[1,2].

소프트웨어 제품의 정확한 인도시기(Release times)를 예측하거나 효용성 및 신뢰성을 예측하기 위해서는 소프트웨어 테스팅 과정에서 중요한 요소인 테스트 커버리지(Coverage)를 이용하면 보다 효율적인 테스팅 작업을 할 수 있다. 이러한 모형은 기존에 존재하는 NHPP모형에서 테스트 커버리지를 포함하는 모형이 된다. 이런 모형을 ENHPP(Enhanced non-homogeneous Poission process) 이라고 한다[3,4].

본 연구에서는 유한 고장 ENHPP 모형에 대하여 수명분포가 지수를 가지는 Goel-Okumoto 모형 및 S-형과 이들의 중첩 및 혼합모형 및 을 제시하여 신뢰성 척도를 추정하고 이를 바탕으로 소프트웨어 신뢰성에 대한 특성을 알아보고자 한다. 본 논문의 2장에서는 관련연구로서 테스트 커버리지, 단순 커버리지 및 제안된 중첩 및 혼합 커버리지 함수에 에 대하여 서술하였고 3장에서는 ENHPP에 모수추정방법에 대하여 나열하였으며 4장에서는 실제 고장자료를 이용하여 각 모형에 대한 소프트웨어 고장 자료를 분석하였으며 마지막으로 5장에서는 결론을 나열 하였다

Ⅱ. 관련 연구

2.1 소프트웨어 신뢰성에 대한 테스트 커버리지

테스트 커버리지를 유한고장 비동질적 포아송 과정 NHPP에 적용 시킬 수 있다. 이러한 적용 모형을 확장된 비동질적 포아송 과정(ENHPP)이라고 부른다. 이러한 ENHPP 모형의 오류탐색비율은 숨겨진 잠재 오류 사이트의비와 남아있는 오류함의 기대수와의 곱에 비례한다. 이러한 ENHPP 모형은 다음과 같은 가정을 한다[3].

(가정 1) 오류들은 모든 잠재 오류 사이트에 대하여 균일하게(uniformly) 분포되어 있다.

(가정 2) 잠재 오류 사이트가 숨겨졌을 때 사이트에서 오류가 탐색되어 질수 있는 확률은 $K(t)$.

(가정 3) 수리는 효과적으로 즉시 이루어지고 새로운 오류는 발생되지 않는다. (이 가정은 유한 고장 NHPP 모형과 유사하다)

(가정 4) 커버리지는 테스팅 시간에 대하여 연속적인 단조 비 감소함수(Monotonic non-decreasing function)를 따른다.

이러한 모형을 해석학적으로 표시하면 다음과 같이 표현된다.

$$\frac{dm(t)}{dt} = \tilde{a}K(t)\frac{dc(t)}{dt} \quad \cdots\cdots (2.1)$$

혹은 $$m(t) = \tilde{a}\int_0^t K(t)c'(\tau)d\tau.$$

단, $\tilde{a}$ 는 무한 테스팅 시간이 주어졌을 때 탐색될 수 있는 오류의 기대수이고 완전한 오류 탐색 커버리지는 $K(t)=1$이고 완전 테스트 커버리지는 다음을 만족한다.

$$\lim_{t\to\infty} c(t) = 1 \quad \cdots\cdots (2.2)$$

만약 $K(t)$가 상수 K로 가정하면 t시점에서 탐색 될 오류의 기대 수 $m(t)$는 다음과 같이 표현 가능하다.

$$m(t) = \tilde{a}Kc(t) \quad \cdots\cdots (2.3)$$

(2.3)식은 t시점에서 탐색 될 오류의 기대 수는 오류의 총 오류의 기대수와 t시점에서의 오류 탐색 커버리지의 곱과 같다. 따라서 $a=\tilde{a}K$로 하면 탐색 될 오류의 기대 수 $m(t)$는 다음과 같이 또 다른 형태로 표현 가능하다.

$$m(t) = a\,c(t) \quad \cdots\cdots (2.4)$$

이런 결과로 인하여 고장 강도함수(Intensity function) $\lambda(t)$ 을 적용시키면 다음과 같다.

$$\frac{dm(t)}{dt} = \lambda(t) = ac'(t) \quad \cdots\cdots (2.5)$$

고장 강도 함수 $\lambda(t)=[a-m(t)]\,h(t)$임이 됨이 알려져 있으므로[1, 3, 4] (2.5)식을 변형하면 다음과 같이 표현할 수도 있다.

$$\lambda(t)=[a-m(t)]\frac{c'(t)}{1-c(t)}=[a-m(t)]\,h(t) \quad \cdots\cdots (2.6)$$

그러므로 한 개의 오류 당 오류발생비율인 위험함수 (Hazard function) $h(t)$는 다음과 같이 유도 된다.

$$h(t)=\frac{c'(t)}{1-c(t)} \quad \cdots\cdots (2.7)$$

따라서 이러한 ENHPP모형은 테스팅이 진행됨에 따라 각각의 오류가 발생할 수 있는 비율은 시간에 따라 변화하기 때문에 시간에 의존하는 고장 발생 비율이라고 할 수 있다. 또, 이러한 위험함수에 해당하는 분포는 t 시점에서 커버리지 함수를 평가할 수 있다. 이러한 ENHPP모형은 (불완전한) 커버리지 함수에 대하여 실제적인 상황($c(\infty)<1$) 을 적용시킬 수 있다. 결과적으로 조건부 신뢰도(Conditional reliability) $R(t\mid s)$는 다음과 같이 알려져 있다[3].

$$R(t\mid s)=e^{-\int_{s}^{s+t}\lambda(\tau)d\tau}=e^{-a[c(s+t)-c(s)]} \quad \cdots\cdots (2.8)$$

단, t는 임무시간(Mission time)이고 s는 마지막 고장 시간이다.

2.2 커버리지 함수를 이용한 NHPP 모형(ENHPP)

2.2.1 지수 커버리지 함수

잘 알려진 Goel-Okumoto 모형[3, 5, 6]은 이 분야에서 기본적인 모형이다. 이러한 모형은 지수 커버리지 함수를 이용한 모형 즉, 오류 당 고장 발생 비율이 상수를 가지는 모형이 된다. 이러한 지수 모형을 NHPP로 접근하면 평균값 함수와 강도 함수는 다음과 같이 알려져 있다.

$$m(t|a,g)=a(1-e^{-gt}) \quad (a>0,\ g>0) \quad \cdots\cdots (2.9)$$

$$\lambda(t|a,g)=age^{-gt} \quad \cdots\cdots (2.10)$$

단, a 는 무한하게 테스팅 시간이 주어졌을 때 관찰된 고장의 기대수이고 g는 오류 당 고장 발생 비율을 의미한다. 따라서 지수모형에 대한 커버리지 함수 $c(t)$와 위험함수 $h(t)$는 다음과 같이 유도 된다.

$$c(t)=1-e^{-gt} \quad \cdots\cdots (2.11)$$

$$h(t)=\frac{c'(t)}{1-c(t)}=g \quad (2.12)$$

2.2.2 S-형 커버리지 함수

S-형 커버리지 함수 모형[3, 7]은 Yamada-Ohba-Osaki 모형으로 알려져 있고 S-형 모형을 NHPP로 접근하면 평균값 함수와 강도 함수는 다음과 같이 알려져 있다.

$$m(t|a,g)=a\left[1-(1+gt)e^{-gt}\right] \quad (a>0,\ g>0) \quad \cdots (2.13)$$

$$\lambda(t|a,g)=ag^2te^{-gt} \quad \cdots\cdots (2.14)$$

따라서 S-형 모형에 대한 커버리지 함수 $c(t)$와 위험함수 $h(t)$는 다음과 같이 유도 된다.

$$c(t)=1-(1+gt)e^{-gt} \quad \cdots\cdots (2.15)$$

$$h(t)=\frac{c'(t)}{1-c(t)}=\frac{g^2t}{1+gt} \quad \cdots\cdots (2.16)$$

2.3 제안된 중첩 및 혼합(변형) 커버리지 함수

소프트웨어 시스템이 복잡해지면 고장의 원인이 하나의 강도 함수에 의해서만 일어나지 않고 여러 원인이 중첩되어 발생할 수 있다. 이런 경우에 강도함수와 평균값함수를 알고 있으면 모형화가 가능하다. 즉, 어떤 부분시스템(Subsystem)의 고장이 시스템의 고장이 원인(독립적으로)이 되면 중첩모형을 적용하여 시스템 고장을 반영할 수 있다. $N_j(t)$을 시간 (0, t] 사이에 강도함수가 $\lambda_j(t)$ 을 가진 j번째 요소로부터 고장이 발생되는 NHPP라고 표현하면 $N(t)=\sum_{j=1}^{J}N_j(t)\,(j=1,2,\cdots,J$)는 (0, t]의 구간에서 발생된 고장의 총수는 다음과 같은 강도함수와 평균값함수를 가진 중첩 NHPP가 된다[8].

$$\lambda(t)=\lambda_1(t)+\lambda_2(t)+\cdots+\lambda_J(t)) \quad \cdots\cdots (2.17)$$
$$m(t)=m_1(t)+m_2(t)+\cdots+m_J(t)$$

따라서 지수 형과 S-형 커버리지 패턴에 대한 평균값 함수와 강도 함수는 다음과 같이 표현 할 수 있다.

$$\lambda(t)=\lambda_1(t)+\lambda_2(t)=a\left(g_1 e^{-g_1 t}+g_2^2 t\, e^{-g_2 t}\right) \quad \cdots\cdots (2.18)$$

$$\begin{aligned} m(t) &= m_1(t)+m_2(t) \\ &= a\left(2-e^{-g_1 t}-(1+g_2 t)\, e^{-g_2 t}\right) \end{aligned} \quad \cdots\cdots (2.19)$$

단, $m_1(t)=a(1-e^{-g_1 t})$, $\lambda_1(t)=a g_1 e^{-g_1 t}$

$m_2(t)=a\left[1-(1+g_2 t)e^{-g_2 t}\right]$,

$\lambda_2(t)=a g_2^2\, t\, e^{-g_2 t}$. $a>0,\ g_1>0,\ g_2>0$.

즉, 중첩에 대한 커버리지 함수 $c(t)$와 위험함수 $h(t)$는 다음과 같이 유도 된다.

$$c(t)=2-e^{g_1 t}-(1+g_2 t)\, e^{-g_2 t} \quad \cdots\cdots (2.20)$$

$$h(t)=\frac{g_1 e^{-g_1 t}+g_2^2 t e^{-g_2 t}}{e^{-g_1 t}+(1+g_2 t)\, e^{-g_2 t}-1} \quad \cdots\cdots (2.21)$$

한편, 부품이나 시스템집단의 수명분포가 서로 다른 k개의 부분집단으로 구성된 경우에 적용 가능한 모형이 혼합모형이다[9]. 예를 들어 전체집단이 결합부품이나 제조상의 결함을 지닌 초기고장 집단과 정상집단의 두 그룹으로 구성되는 경우도 혼합모형에 해당된다. 이러한 혼합모형은 분포함수와 확률밀도함수는 여러분포의 가중 합 형태로 다음과 같이 표현 된다

$$F(t|\Theta,p)=\Pr(T<t)=\sum_{i=1}^{k} p_i F_i(t|\theta_i) \quad \cdots\cdots (2.22)$$

단, $\sum_{i=1}^{k} p_i = 1$

따라서 지수 형과 S-형 커버리지 패턴에 대한 평균값 함수와 강도 함수의 혼합모형은 다음과 같이 표현 할 수 있다.

$$\begin{aligned} \lambda(t) &= p\,\lambda_1(t)+(1-p)\lambda_2(t) \\ &= a\left(p g_1 e^{-g_1 t}+(1-p) g_2^2\, t\, e^{-g_2 t}\right) \end{aligned} \quad \cdots\cdots (2.23)$$

$$m(t)=p m_1(t)+(1-p)\, m_2(t)$$

$$=ap(1-e^{-g_1 t})+a(1-p)\left[1-(1+g_2 t)e^{-g_2 t}\right] \quad (2.24)$$

단, $a>0,\ p\geq 0,\ g_1>0,\ g_2>0$.

즉, 혼합에 대한 커버리지 함수 $c(t)$와 위험함수 $h(t)$는 다음과 같이 유도 된다.

$$c(t)=(p-1)(1+g_2 t)\, e^{-g_2 t}-p e^{g_1 t}+1 \quad \cdots\cdots (2.25)$$

$$h(t)=\frac{p g_1 e^{-g_1 t}+(1-p) g_2^2 t\, e^{-g_2 t}}{p\, e^{-g_1 t}+(1-p)(1+g_2 t)e^{-g_2 t}}$$

III. ENHPP 모형에 대한 모수 추정

시간$(0,t]$까지 조사하기 위한 시간 절단(Time truncated)모형은 n번째 까지 고장시점 자료를

$$x_k=\sum_{i=1}^{k} t_k \ (k=1,2,\cdots,n\,;0\leq x_1\leq x_2\leq\cdots\leq x_n) \quad \cdots (3.1)$$

이라고 하면 데이터 집합 D_t는 $\{n,x_1,x_2\cdots x_n\,;t\}$와 같이 구성된다. n번째까지 고장시점이 관찰된 고장 절단 모형일 경우에 데이터 집합 D_{x_n}은 $\{x_1,x_2\cdots x_n\}$으로 구성된다. 이 시간 절단 모형에서의 우도함수는 다음과 같이 알려져 있다[1, 3].

$$L(\Theta\mid D_t)=\left(\prod_{i=1}^{n}\lambda(x_i)\right)\exp(-m(t)) \quad \cdots\cdots (3.2)$$

단, Θ은 미지의 모수 집합을 의미하고 우도함수 (3.2)식에서 t을 x_n으로 대치하면 유사한 형태의 고장 절단 모형의 우도함수가 된다. 본 논문에서 이러한 우도함수를 이용한 최우추정법(Maximum Likelihood Estimation; MLE)으로 모수 추정을 실시하고자 한다. 지수 커버리지 모형에 대한 로그 우도 함수 는 (3.2)식에 (2.9)을 (2.10)에 대입하면 다음과 같다[3].

$$L_{\mathrm{Exp}}(a,g\mid D_{x_n})=\left(\prod_{k=1}^{n} a g\, e^{-g x_k}\right)\exp\left[-a(1-e^{-g x_n})\right] \quad \cdots\cdots (3.3)$$

MLE을 이용하기 위한 로그우도함수를 구하면 다음과 같이 유도된다.

$$\ln L_{\mathrm{Exp}} = n\ln a + n\ln g - g\sum_{k=1}^{n} x_k - a\left(1-e^{-g x_n}\right) \quad (3.4)$$

(3.4) 식을 이용하여 각 모수에 대한 최우추정량 $\hat{a}_{MLE}$ 와 $\hat{g}_{MLE}$ 은 다음 식을 만족한다.

$$\frac{n}{\hat{a}} = 1-\exp(-\hat{g})\, x_n \quad \cdots\cdots (3.5)$$

$$\frac{n}{\hat{g}} = \sum_{k=1}^{n} x_k + \hat{a}\, x_n \exp(-\hat{g}\, x_n) \quad \cdots\cdots (3.6)$$

(3.5)와 (3.6)식을 수치적으로 풀어 두 모수를 구할 수 있다. 유사한 방법으로 S-커버리지 모형의 모수에 대한 최우추정량 $\hat{a}_{MLE}$ 와 $\hat{g}_{MLE}$ 은 다음 식을 만족한다.

$$\frac{n}{\hat{a}} = 1-\exp(-\hat{g})-\hat{g}\, x_n \exp(-\hat{g}\, x_n) \quad \cdots\cdots (3.7)$$

$$\frac{2n}{\hat{g}} = \sum_{k=1}^{n} x_k + \hat{a}\ \hat{g}\ x_n^2\ \exp(-\hat{g}\ x_n) \quad \cdots\cdots (3.8)$$

(3.7) 과 (3.8)식을 수치적으로 풀어 두 모수의 MLE 추정량을 구할 수 있다.

유사한 방법으로 (3.2), (2.18) 과 (2.19)를 관련하면 중첩모형에 대한 우도함수는 다음과 같이 표현된다[8].

$$L_{NHPP}(D_{\boldsymbol{s}}) = \left(\prod_{k=1}^{n}(\lambda_1(x_k)+\lambda_2(x_k))\right)\exp[-m_1(t)-m_2(t)] \quad \cdots\cdots (3.9)$$

단, $m_1(t) = a(1-e^{-g_1 t}),\quad \lambda_1(t) = a g_1 e^{-g_1 t}$

$$m_2(t) = a\left[1-(1+g_2 t)e^{-g_2 t}\right],$$

$$\lambda_2(t) = a g_2^2\ t\ e^{-g_2 t}.\quad a>0,\ g_1>0,\ g_2>0.$$

(3.9) 식을 이용하여 MLE을 이용하기 위한 로그우도함수를 구하여 각 모수에 대한 최우추정량 $\hat{a}_{MLE}$, $\hat{g}_{1MLE}$ 와 $\hat{g}_{2MLE}$ 은 다음 식을 만족한다.

$$\frac{n}{\hat{a}} = 2-e^{-\hat{g}_1 x_n}-(1+\hat{g}_2 x_n)\, e^{-\hat{\beta}_2 x_n} \quad \cdots\cdots (3.10)$$

$$\sum_{k=1}^{n}\frac{e^{-\hat{g}_1 x_k}-\hat{g}_1\, x_k\, e^{-\hat{g}_1 x_k}}{\hat{g}_1\ e^{-\hat{g}_1 x_k}+\hat{g}_2^2\ x_k\, e^{-\hat{g}_2 x_k}} = \hat{a}\ x_n\, e^{-\hat{g}_1 x_n} \quad \cdots (3.11)$$

$$\sum_{k=1}^{n}\frac{2\hat{g}_2^2\, x_k\, e^{-\hat{g}_2 x_k}-\hat{g}_2^3\, x_k\, e^{-\hat{g}_2 x_k}}{\hat{g}_1\ e^{-\hat{g}_1 x_k}+\hat{g}_2^2\ x_k\, e^{-\hat{g}_2 x_k}} \quad \cdots\cdots (3.12)$$

$$= \hat{a}\,(1-\hat{g}_2 x_n^2)\, e^{-\hat{g}_2 x_n}$$

유사한 방법으로 (3.2), (2.23) 과 (2.24)를 관련하면 혼합모형에 대한 우도함수는 다음과 같이 표현된다[9].

$$L_{NHPP}(D_{\boldsymbol{s}}) = \left(\prod_{k=1}^{n}(p\lambda_1(x_k)+(1-p)\lambda_2(x_k))\right) \quad \cdots\cdots (3.13)$$
$$\cdot \exp[-(pm_1(t)+(1-p)m_2(t))]$$

단, $m_1(t) = a(1-e^{-g_1 t}),\quad \lambda_1(t) = a g_1 e^{-g_1 t}$

$$m_2(t) = a\left[1-(1+g_2 t)e^{-g_2 t}\right],$$

$$\lambda_2(t) = a g_2^2\ t\ e^{-g_2 t}.\quad a>0,\ g_1>0,\ g_2>0$$

(3.13) 식을 이용하여 MLE을 이용하기 위한 로그우도함수를 구하여 각 모수에 대한 최우추정량 $\hat{a}_{MLE}$, $\hat{p}_{MLE}$, , $\hat{g}_{1MLE}$ 와 $\hat{g}_{2MLE}$ 은 다음 식을 만족한다. 그러나 본 논문에서는 $\hat{p}$을 사전에 미리 알고 있는 상수(정수)라고 가정하여 수식을 전개하고자 한다.

$$\frac{n}{\hat{a}} = (p-1)(1+g_2 x_n)\, e^{-g_2 x_n} - pe^{g_1 x_n} + 1 \quad \cdots\cdots (3.14)$$

$$\sum_{k=1}^{n}\frac{pe^{-\hat{g}_1 x_k}-p\hat{g}_1\, x_k\, e^{-\hat{g}_1 x_k}}{p\hat{g}_1\ e^{-\hat{g}_1 x_k}-p\hat{g}_2^2\ x_k\, e^{-\hat{g}_2 x_k}+g_2^2 x_k e^{-g_2 x_k}} \quad \cdots (3.15)$$

$$= p\hat{a}\, x_n\, e^{-\hat{g}_1 x_n}$$

$$\sum_{k=1}^{n}\frac{pe^{-\hat{g}_1 x_k}-p\hat{g}_1\, x_k\, e^{-\hat{g}_1 x_k}}{p\hat{g}_1\ e^{-\hat{g}_1 x_k}-p\hat{g}_2^2\ x_k\, e^{-\hat{g}_2 x_k}+g_2^2 x_k e^{-g_2 x_k}} \quad \cdots (3.16)$$

$$= \hat{a}\,(1-p)(1-\hat{g}_2 x_n^2)\, e^{-\hat{g}_2 x_n}$$

Ⅳ. 소프트웨어 고장 자료 분석

이 장에서 실제 고장자료를 이용하여 모형을 분석하고자 한다. 이 고장자료는 NTDS(Naval Tactical Data System)에 의해 발생된 소프트웨어 고장 자료로서 Goel과 Okumoto[6], Mazzuchi 와 Soyer[10] 등이 이 고장자료를 이용하여 소프트웨어 모형을 제안 한 바 있다. 본 연구도 이 자료를 이용하고자 한다. 이 자료는 〈표 1〉에 나열 되어 있다. 본 논문에서는 모수추정을 위하여 최우추정법을 이용하였고 비선형 방정식의 계산방법은 수치 해석적 기본 방법인 이분법(Bisection method)을 사용하였다. 이러한 계산은 초기 값을 10^{-5} 와 10 을, 허용 한계(Tolerance for width of interval)는 10^{-10} 을 주고 수렴 성을 확인하면서 충분한 반복 횟수인 100 번을 C-언어를 이용하여 모수 추정을 수행하였다.

표 1. NTDS 자료
Table 1. Failure Interval Time Data

고장번호(i)	고장간격시간(t_i)	누적고장시간(x_i)
1	9	9
2	12	21
3	11	32
4	4	36
5	7	43
6	2	45
7	5	50
8	8	58
9	5	63
10	7	70
11	1	71
12	6	77
13	1	78
14	9	87
15	4	91
16	1	92
17	3	95
18	3	98
19	6	104
20	1	105
21	11	116
22	33	149
23	7	156
24	91	247
25	2	249
26	1	250
27	87	337
28	47	384
29	12	396
30	9	405
31	135	540

Goel-Okumoto 커버리지 모형, S-형 커버리지 모형 중첩 모형 및 p에 의존한 혼합모형에 대한 모수의 추정 값들은 결과는 〈표 2〉에 요약되었다. 이 분야에서는 일반적으로 모형 선택의 하나의 방법으로 편차자승합(SSE [9])을 이용할 수 있다. 이러한 결과는 〈표 3〉에 요약되었다. 이 표에서 Goel-Okumoto 커버리지 모형과 S-형 커버리지 모형에 비해 중첩모형 상대적으로 효율적 모형으로 나타나고 있지만 p값이 상승함에 따라 혼합모형이 더 효율적 모형으로 나타나고 있다. 다.

〈그림 1〉 은 각 모형에 대하여 위험함수를 그린 그림이다. 이 그림에서는 지수 커버리지 모형은 일정하고 S-커버리지 모형과 혼합모형($p=0.5$)은 증가형태를 가지지만 중첩 커버리지 모형은 감소형태를 나타내고 있으며 〈그림 2〉에서는 혼합모형에 대한 위험함수를 그린 그림으로서 p 값이 작을수록 높은 확률 패턴을 가지고 있다.

〈그림 3〉은 각 모형에 대한 커버리지 함수를 그린 그림이다. ENHPP 모형에서 가정한 대로 연속적인 단조 비 감소 함수형태로 나타나고 있다. 〈그림 4〉에서는 혼합모형에 대한 커버리지 함수를 그린 그림으로서 p 값이 작을수록 높은 확률 패턴을 가지고 있다. 〈그림 5〉에서는 신뢰도에 대한 패턴을 알아보았다. 이 그림에서도 역시 중첩 및 혼합모형이 높게 나타나고 있다.

표 2. 각 모형의 모수 추정 값
Table 2. Estimator of each model

추정치 / Model		$\hat{a}$	$\hat{g}_1$	$\hat{g}_2$
Goel-Okumoto 모형		35.07	0.00398	-
Yamada-Ohba-Osaki 모형		34.86	-	0.00684
중첩모형(Superposition)		32.75	0.00542	0.00063
혼합모형 (Mixture)	$p=0.1$	34.89	0.006156	0.0003582
	$p=0.2$	34.08	0.005471	0.0003184
	$p=0.3$	33.07	0.47880	0.0002786
	$p=0.4$	33.01	0.004104	0.0002388
	$p=0.5$	32.07	0.003420	0.0001990
	$p=0.6$	32.01	0.002736	0.0001592
	$p=0.7$	31.09	0.002052	0.0001194
	$p=0.8$	31.04	0.001368	0.0000796
	$p=0.9$	30.07	0.000684	0.0000398

표 3. 모형들에 대한 SSE의 값
Table 3. SSE of each model

Model		*SSE*
Goel-Okumoto 모형		9013.172
Yamada-Ohba-Osaki 모형		5049.463
중첩모형(Superposition)		2078.594
혼합모형 (Mixture)	$p=0.1$	4091.568
	$p=0.2$	3992.456
	$p=0.3$	3082.754
	$p=0.4$	2575.873
	$p=0.5$	2065.876
	$p=0.6$	1995.897
	$p=0.7$	1880.675
	$p=0.8$	1809.345
	$p=0.9$	1802.453

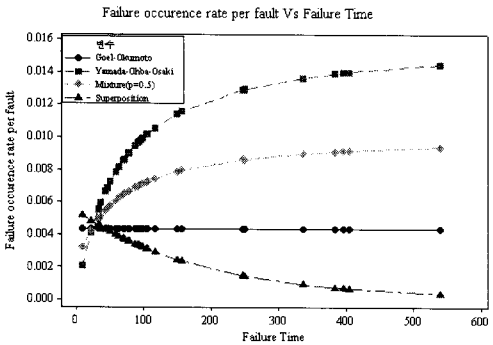

그림 1. 각 모형에 대한 위험함수
Figure 1. Hazard function of each model

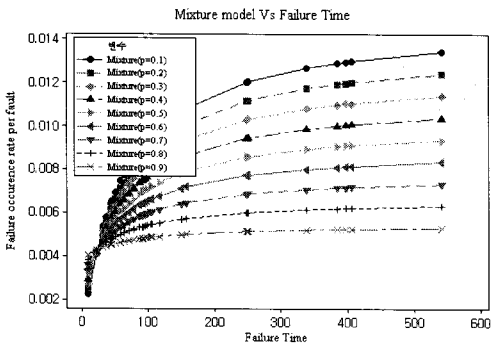

그림 2. 혼합 모형에 대한 위험함수
Figure 2. Hazard function of mixture model

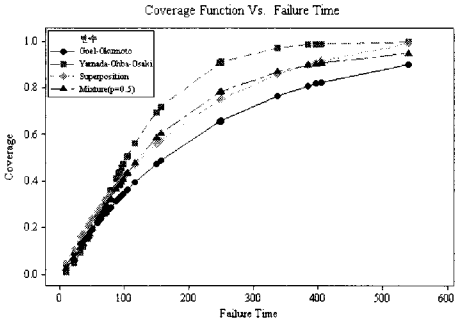

그림 3. 각 모형에 대한 커버리지 함수
Figure 3. Coverage function of each model

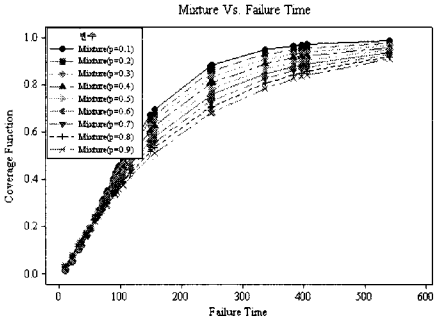

그림 4. 혼합 모형에 대한 커버리지함수
Figure 4 Coverage function of mixture model

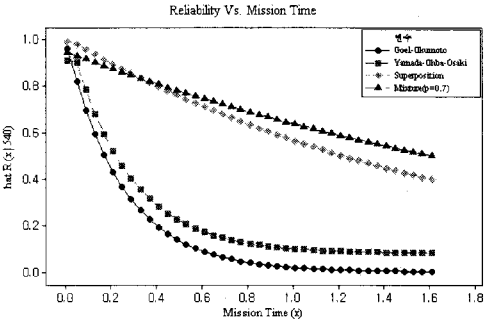

그림 5. 각 모형에 대한 신뢰도
Figure 4 Reliability of each model

V. 결 론

본 논문에서는 기존에 존재하는 NHPP모형에서 테스트 커버리지를 포함하는 모형인 중첩 및 혼합 ENHPP 모형에 대하여 연구하였다. 즉, 잔존 결함 1개당 고장 발생률이 감소추세를 가진 중첩 형태 및 혼합 커버리지 모형을 비교하는 내용을 제안하였다. 고장 간격시간으로 구성된 자료를 이용하여 기존의 모형과 중첩 및 혼합 모형에 대하여 최우추정법을 이용하여 모수 추정을 실시하였다. 수치적인 예에서는 고장수가 비교적 큰 실측 자료인 NTDS 자료를 통하여 분석하였다. 효율적인 모형 비교를 위한 편차자승합의 결과는 중첩 커버리지 모형과 혼합모형이 기존에 잘 알려진 지수 커버리지 모형이나 S-커버리지 모형보다 우수함을 보이고 있다. 따라서 중첩이나 혼합 ENHPP 모형도 이 분야에서 효율적으로 이용 할 수 있는 모형이 됨을 알 수 있었다. 향후 이러한 중첩 분포를 이용한 베이지안적 접근 방법(p을 사전 정보로 활용)이나 수리적인 추정과 검정 부분에 대한 수리적인 접근이 기대 된다.

참고문헌

[1] 김 희 철, 최 유 순, 박 종 구, "어랑 분포를 이용 한 NHPP 소프트웨어 신뢰성장 모형에 관한 연구", 한국 해양 정보 통신 학회 논문 지, 10권1호 pp.7-14, 2006.

[2] 김 희 철 "지수화 지수분포에 의존한 NHPP 소프트웨어 신뢰성장 모형에 관한 연구", 한국 컴퓨터정보학회 논문 지, 11권 5호, pp.9-18, 2006.

[3] S. S. Gokhale and K. S. Trivedi. "A time/structure based software reliability model". Annals of Software Engineering. 8, pp. 85-121. 1999

[4] 김 희 철 "Burr 커버리지 함수에 기초한 ENHPP 소프트웨어 신뢰성장모형에 관한 연구", 한국 컴퓨터정보 학회 논문 지, 12권 4호, pp.33-42, 2007.

[5] J. F. Lawless, "Statistical Models and Methods for Lifetime Data", John Wiley& Sons, New York, 1981.

[6] A. L. Goel and K. Okumoto, "Time-Dependent Error-Detection Rate Models for Software Reliability and Other Performance Measures". IEEE Trans. on Reliability R-28(3): pp.206-211, Aug. 1979.

[7] S. Yamada, M. Ohba and S. Osaki. "S-Shaped Reliability Growth Modeling for Software Error Detection". IEEE Trans. on Reliability. R-32(5): pp. 475-485, Dec. 1983.

[8] E, Cinlar. "Introduction To Stochastic Process", New Jersey: Prentice-Hall, 1975.

[9] D. A. Berry and D. K. Stangl. "Bayesian Bio-statistics ", MARCEL DEKKER INC, pp.601-616, 1996.

[10] T. A., Mazzuchi and R., Soyer. "A Bayes Empirical Model for Software Reliability". IEEE Transactions on Reliability, 37, pp 248-254, 1988

[11] S. S. Gokhale and T. Philip and P. N. Marinos and k. S. Trivedi, "Unification of Finite Failure Non-Homogeneous Poisson Process Models through Test Coverage", In Proc. of Intl. Symposium on software Reliability Engineering ," White Plains. NY, 1996.

[12] J. R. Horgan, S. London, and M.R. Lyu, " Achieving Software Quality with Testing Coverage Measure," IEEE Computer, pp. 60-69, Sept. 1994.

[13] L, Kuo and Yang, T. Y. "Bayesian Computation of Software Reliability". Journal of the American Statistical Association, Vol.91, pp.763-773, 1996.

저 자 소 개

김 희 철

1992년 동국대학교 통계학과 석사
1998년 동국대학교 통계학과 박사
2000년 3월 ~ 2004년 2월
송호대학 정보산업계열 조교수
2005년 3월~현재 남서울대학교
산업경영공학과 전임강사
관심분야 : 소프트웨어 신뢰성공학, 웹프로그래밍,전산통계, 인터넷비즈니스

김 평 구

1984년 고려대학교 통계학과 석사
1994년 동국대학교 통계학과 박사
1990년 3월 ~ 현재
충청대학 품질경영과 교수
관심분야 : 품질경영